

SYSTEM, METHOD AND COMPUTER PROGRAM PRODUCT FOR
CREATING A DESCRIPTION FOR A DOCUMENT OF A REMOTE
NETWORK DATA SOURCE FOR LATER IDENTIFICATION OF THE
DOCUMENT AND IDENTIFYING THE DOCUMENT UTILIZING A
DESCRIPTION

ABSTRACT

A system, method and computer program product are provided for creating a description of a document of a remote network data source for later identification of the document. Information about a document on a remote network data site is received from a user. A document identifier is created based on the user-input information. The document identifier identifies the particular document. A markup language description is retrieved. The markup language description defines properties of elements of a document in a markup language. The document and the content of the document are analyzed utilizing the document identifier and the markup language description. A description of the document is generated based on the analysis. The document description is stored. A system, method and computer program product are also provided for identifying a document. A document is received. Document descriptions of several documents are also received. The document descriptions are compared with the document. A document recognition score is calculated for each of the document descriptions based on a likelihood that the document description matches the document. A document description is selected based at least in part on the document recognition scores. The document is identified based on the selected document description. A system, method and computer program product are provided for identifying documents. A document is analyzed. A description of the document is created based on the

analysis. The document is recognized utilizing the document description. A determination is made as to whether the document is in a list of pre-identified documents.

FOR OFFICIAL USE ONLY